

ARTIKEL
PENELITIAN DISERTASI DOKTOR

**TES KEMAMPUAN BERPIKIR TINGKAT TINGGI
FISIKA DI SMA LANGKAH PENGEMBANGAN
DAN KARAKTERISTIKNYA**



Oleh:

Drs. EDI ISTIYONO, M.Si.

UNIVERSITAS NEGERI YOGYAKARTA
NOVEMBER 2013

Dibiayai oleh DIPA Universitas Negeri Yogyakarta
dengan Surat Perjanjian Penugasan dalam rangka Pelaksanaan Program
Penelitian Disertasi Doktor Tahun Anggaran 2013
Nomor: 532a/BOPTN/UN34.21/2013 Tanggal 27 Mei 2013

TES KEMAMPUAN BERPIKIR TINGKAT TINGGI FISIKA DI SMA LANGKAH PENGEMBANGAN DAN KARAKTERISTIKNYA

Oleh:

Edi Istiyono

Jurusan Pendidikan Fisika FMIPA Universitas Negeri Yogyakarta

edi_istiyono@uny.ac.id dan edi_istiyono_uny@yahoo.co.id

Abstrak

Penelitian ini bertujuan: (1) mengembangkan instrumen untuk mengukur kemampuan berpikir tingkat tinggi dalam mata pelajaran Fisika (PhysTHOTS) di SMA, dan (2) mendapatkan karakteristik instrumen PhysTHOTS. Penelitian ini dilakukan melalui dua tahap, yakni: pengembangan awal instrumen dan ujicoba. Pengembangan awal instrumen meliputi penyusunan draf yang berupa: *learning continuum*, kisi-kisi instrumen, item-item, dan pedoman penilaian. Draft tersebut divalidasi oleh ahli pengukuran, ahli pendidikan fisika, ahli fisika, dan praktisi. Instrumen terdiri atas dua perangkat tes yang telah divalidasi diujicobakan pada 1001 siswa. Instrumen terdiri atas dua perangkat tes yang masing-masing dikemas untuk uji coba yang masing-masing perangkat tes memuat 30% *anchor item* untuk *equating*. Uji coba dilakukan pada 1001 siswa di 10 SMAN di Provinsi Daerah Istimewa Yogyakarta. Penskoran dengan skala politomus lima kategori dan dianalisis menggunakan model *Partial Credit* 1-PL. Karakteristik tes terdiri atas kecocokan item, reliabilitas, tingkat kesukaran item, kurva karakteristik item (ICC), fungsi informasi, dan *standard error measurement* (SEM). Untuk uji kecocokan (*goodness of fit*) dan estimasi reliabilitas digunakan analisis dengan Program QUEST. Untuk mendapatkan ICC, fungsi informasi, dan SEM dianalisis dengan multilog. Penskoran politomus dianalisis menggunakan model Rasch melalui program QUEST dan Multilog. Hasil uji coba menunjukkan, dengan kriteria mean dan simpangan baku INFIT MNSQ 1,0 dan 0,0 tes terbukti *fit* dengan model *Partial Credit* 1-PL. Berdasarkan hasil analisis reliabilitas instrumen tersebut diestimasi sebesar 0,95. Berpedoman kriteria batas terendah dan tertinggi INFIT MNSQ sebesar 0,77 dan 1,30 sebanyak 44 item tes *fit* dengan model *Partial Credit* 1-PL. Dari hasil output QUEST dihasilkan indeks kesukaran item rerata (*difficulty*) antara -0,86 sampai 1,06 yang berarti item dalam kategori baik karena nilainya diantara -2,0 dan 2,0. Jadi, instrumen sebanyak 44 item yang diujicobakan semuanya *fit* dengan model dan kategori baik, sehingga dapat digunakan untuk melakukan pengukuran.

Kata kunci: tes, kemampuan berpikir tingkat tinggi, politomos, dan Fisika

PENDAHULUAN

Dewasa ini dunia berada pada era globalisasi. Pada era ini persaingan cukup ketat. Agar dapat menang dalam bersaing, maka negara-negara berlomba-lomba meningkatkan kualitasnya. Kualitas bangsa ditentukan dengan tingkat pendidikan bangsa tersebut. Untuk meningkatkan kualitas pendidikan dapat dimulai dari meningkatkan kualitas pembelajaran. Untuk meningkatkan kualitas pembelajaran dapat dimulai dengan menyusun tujuan pembelajaran yang tepat. Salah satu tujuan Mata Pelajaran Fisika di SMA agar peserta

didik memiliki kemampuan mengembangkan kemampuan bernalar dalam berpikir analisis induktif dan deduktif dengan menggunakan konsep dan prinsip fisika untuk menjelaskan berbagai peristiwa alam dan menyelesaikan masalah baik secara kualitatif maupun kuantitatif (BSNP, 2006: 160). Dengan demikian, melalui pembelajaran Fisika diharapkan peserta didik dapat mengembangkan diri dalam berpikir. Peserta didik dituntut tidak hanya memiliki kemampuan berpikir tingkat rendah (*lower order thinking*), tetapi sampai pada kemampuan berpikir tingkat tinggi (*higher order thinking, HOT*).

Berkenaan dengan kemampuan berpikir tingkat tinggi ini, fakta menunjukkan bahwa prestasi fisika yang diukur pada aspek *reasoning* Indonesia berada pada ranking 40 dari 42 negara (TIMSS & PIRLS International Study Center, 2012:48). Hal senada dinyatakan Ridwan Efendi (2011: 393) bahwa berdasarkan hasil TIMSS dapat ditarik kesimpulan sebagai berikut: (1) rata-rata capaian fisika siswa Indonesia ditinjau dari aspek kognitif (*knowing, applying, reasoning*) masih rendah; (2) kecenderungan capaian fisika siswa Indonesia selalu menurun pada tiap aspek kognitif sehingga kemampuan fisika siswa Indonesia harus ditingkatkan pada semua aspek, khususnya aspek *reasoning* dengan cara membekalkan pada siswa kemampuan berpikir tingkat tinggi. Dengan demikian, prestasi fisika yang menuntut kemampuan berpikir tingkat tinggi siswa sekolah menengah Indonesia, di kancah internasional masih rendah. Prestasi belajar fisika rendah dapat disebabkan karena proses pembelajaran atau model asesmennya yang tidak tepat. Dalam hal ini hanya akan dibahas tentang asesmennya, karena dengan asesmen yang tepat dapat mendorong siswa untuk belajar dengan berpikir tingkat tinggi.

Berdasarkan teori perkembangan Piaget, tahap operasional formal adalah tahap anak mulai dari sebelas tahun. Pada tahap ini anak sudah mulai dapat mengembangkan kemampuan untuk memanipulasi konsep abstrak melalui penggunaan proposisi dan hipotesis (Reedal, 2010:7). Usia Siswa SMA antara 15 sampai 18 tahun, sehingga kemampuan berpikir tingkat tinggi siswa SMA sudah mapan.

Kalau Piaget mengatakan harus menunggu kematangan dan kesiapan seseorang serta harus cocok antara pengaruh dari luar dan perkembangan di dalam dirinya (*match*), tetapi Vygotsky tidak. Ada sesuatu di atas tahap perkembangan itu (*plus one matching*). Ada daerah-daerah yang sangat sensitif untuk diaktualisasikan dalam diri anak yang dinamakan *Zone Proximal Development (ZPD)*. Dengan menerapkan konsep ZPD pada pendidikan, maka pembelajaran akan memajukan perkembangan anak. Salah satu wujud

konkrit implikasi dari teori Vygotsky adalah dilaksanakannya akselerasi belajar bagi anak berbakat, pendidikan bagi anak-anak yang mempunyai kemampuan intelektual luar biasa dan dalam proses pembelajaran harus selalu meningkatkan kadar mental atau berpikir tingkat tinggi.

Menurut Bloom yang telah direvisi terbagi menjadi kemampuan berpikir tingkat rendah (*Lower Order Thinking*) dan kemampuan berpikir tingkat tinggi (*Higher Order Thinking*). Kemampuan yang termasuk *LOT* adalah kemampuan mengingat (*remember*), memahami (*understand*), dan menerapkan (*apply*), sedangkan *HOT* meliputi kemampuan menganalisis (*analyze*), mengevaluasi (*evaluate*), dan menciptakan (*create*) (Anderson and Krathwohl, 2001:30). Taksonmi Bloom sudah lama diterapkan dalam bidang pendidikan dan sudah lama digunakan. Taksonomi Bloom masih digunakan dalam banyak kurikulum dan bahan pengajaran (Brookhart, 2010: 39, Schraw and Robinson, 2011: 158-159). Dengan demikian kemampuan berpikir tingkat tinggi Fisika (*Physics Higher Order Thinking*) meliputi kemampuan fisika dalam menganalisis, mengevaluasi, dan menciptakan.

Untuk memantau proses, kemajuan, dan perbaikan hasil belajar peserta didik secara berkesinambungan, diperlukan penilaian. Penilaian pendidikan adalah proses pengumpulan dan pengolahan informasi untuk menentukan pencapaian hasil belajar peserta didik (Peraturan Menteri Pendidikan Nasional No 20, 2007). Penilaian dapat dilakukan secara lisan maupun tertulis. Penilaian secara tertulis dilakukan dengan tes tertulis. Secara garis besar ada dua bentuk soal tes tertulis, yaitu: memilih jawaban dan mensuplai jawaban. Soal tes tertulis yang jawabannya dengan memilih jawaban antara lain: pilihan ganda, dua pilihan (benar-salah, ya-tidak), menjodohkan, dan sebab-akibat.

Kenyataan bahwa tes pilihan ganda lebih banyak digunakan dari pada bentuk tes yang lain. Hal ini karena tes pilihan ganda memiliki kelebihan-kelebihan, antara lain: (1) materi yang diujikan dapat mencakup sebagian besar bahan pembelajaran, (2) jawaban siswa dapat dikoreksi dengan mudah dan cepat, (3) jawaban setiap pertanyaan sudah pasti benar atau salah, sehingga penilaian objektif (Nana Sujana, 1990:49). Walaupun ada juga kelemahan tes ini, yaitu: (1) kemungkinan peserta didik untuk melakukan tebakan jawaban masih cukup besar dan (2) proses berpikir siswa tidak dapat dilihat dengan nyata (Nana Sujana, 1990:49).

Berdasarkan hasil survei pendahuluan dengan cara melakukan wawancara dengan guru-guru Fisika SMA di Daerah Istimewa Yogyakarta (DIY), sebagian besar di sekolah,

baik pada tes tengah semester maupun tes akhir semester umumnya menggunakan tes pilihan ganda biasa. Jadi, tes pilihan ganda masih merupakan primadona dalam mendapatkan data prestasi belajar Fisika siswa SMA. Kenyataan bahwa tes pilihan ganda yang digunakan di SMA untuk tes hasil belajar mata pelajaran Fisika baru mengukur kemampuan: mengingat, memahami, dan menerapkan. Jadi tes pilihan ganda yang digunakan di SMA baru mengukur kemampuan berpikir tingkat rendah (*Lower Order Thinking (LOT)*) belum mengukur kemampuan berpikir tingkat tinggi Fisika (*PhysHOT*).

Berdasarkan hasil survei pendahuluan juga bahwa penskoran hasil tes dengan model dikotomis, artinya jika item benar diberi skor 1 dan jika salah diberi skor 0. Belum menggunakan model politomis yang lebih adil karena mempertimbangkan langkah-langkah penyelesaian tes. Dengan model penskoran dikotomis ini belum menghargai tahap-tahap penyelesaian soal, karena dengan tingkat kesalahan yang berbeda mendapatkan skor yang sama yakni 0. Dengan demikian skoring model ini tentu kurang adil.

Berdasarkan uraian di atas, untuk mengukur kemampuan berpikir tingkat tinggi Fisika digunakan tes berbentuk pilihan ganda beralasan yang dinamakan Tes kemampuan berpikir tingkat tinggi fisika (*Physics Test for Higher Order Thinking Skills (PhysTHOTS)*). Untuk itu perlu disusun instrumen penilaian kemampuan berpikir tingkat tinggi Fisika (*PhysTHOTS*) yang terdiri atas tes dan pedoman penilaian. Dengan demikian diperlukan adanya penelitian pengembangan instrumen penilaian kemampuan berpikir tingkat tinggi fisika yang terdiri atas tes berpikir tingkat tinggi Fisika (*Physics Test for Higher Order Thinking Skills (PhysTHOTS)*) dan pedoman penilaian.

Masalah utama yang akan diangkat dalam penelitian ini adalah: “Bagaimanakah menyusun instrumen untuk mengukur kemampuan berpikir tingkat tinggi dalam mata pelajaran Fisika (*PhysTHOTS*) di SMA kelas XI pada setiap aspeknya yang disertai karakteristiknya yang meliputi kemampuan menganalisis, mengevaluasi, dan menciptakan?”

Sejalan dengan rumusan masalah yang akan diselesaikan, maka tujuan penelitian ini adalah (1) menghasilkan instrumen untuk mengukur kemampuan berpikir tingkat tinggi dalam mata pelajaran Fisika, dan (2) mendapatkan karakteristik instrumen penilaian

kemampuan berpikir tingkat tinggi dalam mata pelajaran Fisika di SMA kelas XI yang meliputi kemampuan menganalisis, mengevaluasi, dan menciptakan

METODE PENELITIAN

Langkah-langkah pengembangan instrumen berupa tes menggunakan modifikasi Model Wilson dan Model Oriondo dan Antonio, yakni: (1) perancangan tes dan (2) uji coba tes.

Tahap perancangan tes meliputi: (1) Penentuan Tujuan Tes, (2) Penentuan Kompetensi yang akan Diujikan, (3) Penentuan Materi yang Diujikan, (4) Penyusunan Kisi-kisi Tes, (5) Penulisan Item Berdasarkan Prinsip-prinsip Pengembangan Tes *HOT*, (6) Validasi Item Tes, (7) Perbaikan Item dan Perakitan Tes, dan (8) Penyusunan Pedoman Penskoran. Adapun tahap uji coba meliputi: (1) Penetapan Subjek Uji Coba (SMA), (2) pelaksanaan uji coba.

Berkaitan jumlah sampel, untuk analisis secara *IRT* beberapa ahli pengukuran sebaiknya 200 sampai dengan 1000 orang (Seon, 2009: 3). Untuk analisis dengan Rasch, sampel yang digunakan antara 30 sampai 300 orang (Bond and Fox, 2007: 43; Keeves & Masters, 1999: 12-13). Reckase (2000) menyimpulkan bahwa sampel ukuran minimum yang baik untuk memperkirakan tiga parameter yang meliputi: diskriminasi (daya beda), tingkat kesulitan, dan *pseudoguessing* adalah 300 (Haladyna, 2004:206). Jadi dengan model PCM 1PL peserta didik yang dijadikan subjek coba sebanyak 500 orang sudah lebih dari cukup.

Untuk analisis data Penelitian ini menggunakan *Partial Credit Model* 1 PL (PCM 1-PL) untuk pengujian *fit* item tes kemampuan berpikir tingkat tinggi untuk mata pelajaran Fisika SMA. Dasar pertimbangan yang digunakan, yang pertama bahwa PCM sebagai perluasan RM yang merupakan model 1-PL, dapat menggunakan sampel yang tidak sebesar kalau melakukan kalibrasi data politomus menggunakan model 2-PL atau 3-PL (Keeves & Masters, 1999: 12-13). Kedua, bahwa karakteristik respons terhadap item kemampuan berpikir tingkat tinggi mengikuti PCM.

Untuk analisis data terdiri atas beberapa aspek, yakni: (1) Kecocokan Item Instrumen, (2) Reliabilitas, (3) Kurva Karakteristik Item (ICC), (4) Indeks Kesukaran (b), dan (5) Fungsi Informasi dan SEM.

Pengujian *goodness of fit* untuk tes secara keseluruhan dan testi (*case/person*) secara keseluruhan dikembangkan Adam dan Khoo (1996:30) berdasarkan nilai rerata INFIT *Mean of Square (Mean INFITMNSQ)* beserta simpangan bakunya atau mengamati nilai rata INFIT *t (Mean INFIT t)* beserta simpangan bakunya. Jika rerata UNINFIT MNSQ sekitar 1 dan simpangan bakunya 0,0 atau rerata INFIT *t* mendekati 0 dan simpangan bakunya 1,0, maka keseluruhan tes fit dengan model.

Pengujian Validitas untuk Mengetahui *Fit* Item dan Testi terhadap Model mengikuti kaidah bahwa *Item characteritic curve* (ICC) akan mendatar (*flat*) bila besarnya INFIT MNSQ untuk item atau *e* lebih besar dari satuan logit $> 1,30$ atau $< 0,77$. Akibatnya membentuk *platokurtic curve* dan tidak lagi membentuk *leptokurtic curve* (Keeves & Alagumalai 1999: 36). Oleh karena itu, dalam program QUEST ditetapkan bahwa suatu item atau testi/*case/person* dinyatakan *fit* dengan model dengan batas kisaran INFIT MNSQ dari 0,77 sampai 1,30 (Adam & Khoo, 1996:30 & 90). Dalam hal ini menggunakan kisaran nilai *t* adalah ± 2 (pembulatan $\pm 1,96$) jika taraf kesalahan atau *alpha* sebesar 5% (Keeves & Alagumalai 1999: 34-36; Bond & Fox, 2007: 43).

Selain untuk menguji kecocokan, output program QUEST juga menampilkan estimasi reliabilitas set instrumen tes serta indeks kesukarannya. Item dikatakan baik jika indeks kesukaran lebih dari -2.0 atau kurang dari 2.0 Untuk mendapatkan kurva karakteristik item (ICC), fungsi informasi dan SEM digunakan program Multilog. Berdasarkan fungsi informasi dan SEM, maka tes ini cocok untuk siswa dengan kemampuan (θ) rendah, sedang, atau tinggi.

HASIL DAN PEMBAHASAN

Penyiapan Tes

Setelah rumusan *learning continuum* kemampuan berpikir tingkat tinggi Fiaika materi kelas XI SMA, kisi-kisi instrumen, item-iten dan pedoman penilaian direview dan divalidasi oleh ahli pengukuran, ahli Fisika, dan ahli Pendidikan Fisika, dilakukan revisi dan penyempurnaan item-item. Masukan dari forum ahli yaitu untuk menyempurnakan item, terutama dari segi panjang kalimat baik pada soal maupun dalam pedoman penskoran suatu item, penggantian beberapa istilah, dan panampilan grafik atau gambar agar lebih komunikatif.

Tes terdiri atas dua set yakni set I berkode A dan set II berkode B. Setiap tes meliputi materi: gerak analisis vektor, gaya dan getaran, usaha energi, impuls dan momentum yang meliputi aspek dan sub aspek menganalisis, mengevaluasi, dan menciptakan dengan sebaran sebagaimana Tabel 1. Kedua perangkat tes tersebut memiliki delapan item sebagai *anchor item* di dalamnya. Selanjutnya tes dinyatakan layak digunakan dalam oleh beberapa ahli tersebut.

Tabel 1. Sebaran Item Instrumen Kemampuan Berpikir Tingkat Tinggi dalam Mata Pelajaran Fisika di SMA Kelas XI

Aspek	Sub aspek	Materi Fisika SMA Kelas XI			
		Gerak	Gaya	Usaha dan Energi	Momentun dan Impuls
Menganalisis (<i>Analyze</i>)	<i>Differentiating</i>	2 (1A, 9B)	2 (4A, 12B)		1 (21A, 3B)*
	<i>Organizing</i>	1* (19A, 1B)	2 (5A, 13B)	2 (7A, 15B)	2 (9A, 17B)
	<i>Attributing</i>	2 (2A, 10B)	2 (6A, 14B)	1 (20A, 2B) *)	
Mengevaluasi (<i>Evaluate</i>)	<i>Checking</i>		2 (3A, 11B)	2 (8A, 16B)	1 (26A, 8B *)
	<i>Critiquing</i>	2 (11A, 19B)		1 (25A, 7B) *)	2 (10A, 18B)
Menciptakan (<i>Create</i>)	<i>Generating</i>	1*(22A, 4B)	4 (13A, 14A, 12B, 22B)	2 (15A, 23B)	
	<i>Planning</i>	2 (12A, 20B)	1 (23A, 5B) *)		2 (18A, 26B)
	<i>Producing</i>		1 (24A, 6B) *)	2 (16A, 24B)	2 (17A, 25B)

Keterangan:

*) *anchor item* (item sama pada set A dan B)

Uji Coba

Untuk kegiatan uji coba ini terlebih dahulu menentukan SMA di Provinsi DIY yang digunakan untuk uji coba berdasar ranking sekolah berdasar nilai UN Fisika dan kefavoritan di kota /kabupaten. SMA yang digunakan untuk uji coba antara lain: SMA 5 Yogyakarta, SMA 11 Yogyakarta, SMA 1 Bambanglipuro, SMA 1 Sedayu, SMA 1 Wates, SMA 1 Pengasih, SMA 1 Gamping, SMA 1 Minggir, SMA 1 Wonosari, dan SMA 1 Patuk.

Hasil Uji coba

1. Kecocokan Item Instrumen (*goodness fit*)

Pengujian *goodness of fit* untuk tes secara keseluruhan maupun tiap item dengan program Quest. Pengujian Fit tes keseluruhan dikembangkan Adam dan Khoo (1996:30) berdasarkan nilai rerata INFIT *Mean of Square (Mean INFITMNSQ)* beserta simpangan bakunya atau mengamati nilai rata INFIT *t (Mean INFIT t)* beserta simpangan bakunya. Jika rerata UNINFIT MNSQ sekitar 1 dan simpangan bakunya 0,0 atau rerata INFIT *t* mendekati 0 dan simpangan bakunya 1,0, maka keseluruhan tes fit dengan model PCM 1 PL. Berdasarkan Tabel 2, nilai rerata *INFITMNSQ* 1,01 (sekitar 1) dan simpangan baku 0,02 (sekitar 0,0), maka keseluruhan tes fit dengan model PCM 1 PL

Tabel 2. Hasil Estimasi Item Kemampuan berpikir tingkat tinggi Mata Pelajaran Fisika SMA dan Estimasi Testi menurut PCM 1-PL Berdasarkan Data Politomus Empat Kategori

No	Uraian	Estimasi untuk item	Estimasi untuk testi
1	Nilai rata-rata dan simpangan baku	-0,28 ± 0,42	-0,21 ± 0,03
2	Nilai rata-rata dan simpangan baku yang sudah disesuaikan	0,00 ± 0,41	-0,11 ± 0,00
3	Reliabilitas	0,94	0,00
4	Nilai rata-rata dan simpangan baku INFIT MNSQ	1,01 ± 0,02	1,02 ± 0,06
5	Nilai rata-rata dan simpangan baku OUTFIT MNSQ	1,01 ± 0,02	1,01 ± 0,09
6	Nilai rata-rata dan simpangan baku INFIT <i>t</i>	0,85 ± 1,34	0,28 ± 0,92
7	Nilai rata-rata dan simpangan baku OUTFIT <i>t</i>	0,34 ± 0,61	0,11 ± 0,50
8	Item atau testi skor 0	0	0
9	Item atau testi skor perfect	0	0

Pengujian penetapan fit setiap item terhadap model mengikuti kaidah Adam dan Khoo (1996:30), suatu item fit terhadap model jika nilai INFIT MNSQ antara 0,77 smp 1,30. Dengan batas penerimaan item menggunakan INFIT MNSQ atau fit menurut model (antara 0,77 sampai dengan 1,30) dan menggunakan INFIT *t* dengan batas -2,0 sampai 2,0, maka diperoleh item-item yang cocok memenuhi *goodness of fit*. Nilai INFIT MNSQ antara 0,98 sampai dengan 1,05. Dengan batas penerimaan item menggunakan

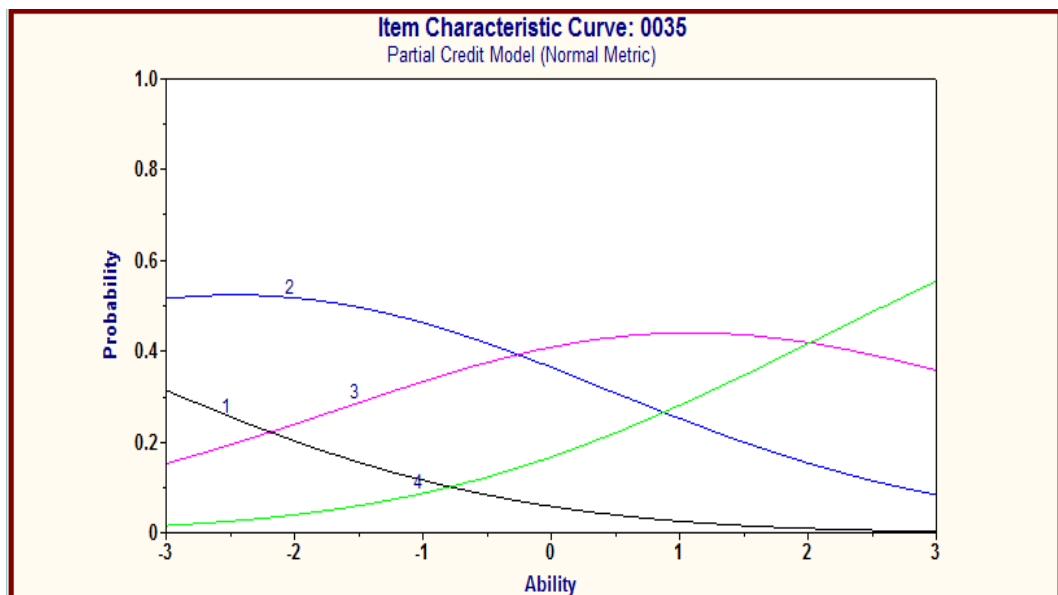
INFIT MNSQ atau fit menurut model (antara 0,77 sampai dengan 1,30) dan menggunakan INFIT t dengan batas $\pm 2,0$, maka semua item sebanyak 44 item *fit* semua.

2. Reliabilitas

Selain untuk menguji kecocokan, output program QUEST juga menampilkan estimasi reliabilitas set instrumen tes. Berdasarkan hasil analisis dengan program QUEST juga didapatkan reliabilitas set instrumen (tes) tersebut diestimasikan sebesar 0,95 yang dinyatakan pada Tabel 2. Nilai reliabilitas ini tergolong kategori tinggi.

3. Kurva Karakteristik Item (*Item Characteristic Curve, ICC*)

Karakteristik item ditunjukkan dengan kurva karakteritik item (ICC) dan indeks kesukaran. Untuk mendapatkan kurva karakteristik item (ICC) digunakan program Multilog. Berdasarkan analisis dengan Multilog diperoleh kurva karakteristik item (ICC) sebanyak 44 buah. Pada Gambar 1 disajikan contoh ICC untuk item 35 (item 17 B), yang dapat dijelaskan bahwa: (a) skor 1(kategori 1) sebagian besar diperoleh siswa dengan ability sangat rendah ($\theta = -3$), (b) skor 2 (kategori 2) sebagian besar diperoleh siswa dengan ability rendah ($\theta = -2$), (c) skor 3 (kategori 3) sebagian besar diperoleh siswa dengan ability tinggi ($\theta = 1,5$), dan (d) skor 4 (kategori 4) sebagian besar diperoleh siswa dengan ability sangat tinggi ($\theta = 3$).



Gambar 1.
Kurva Karakteristik Butir 35 (Butir 17B)

4. Indeks Kesukaran (b)

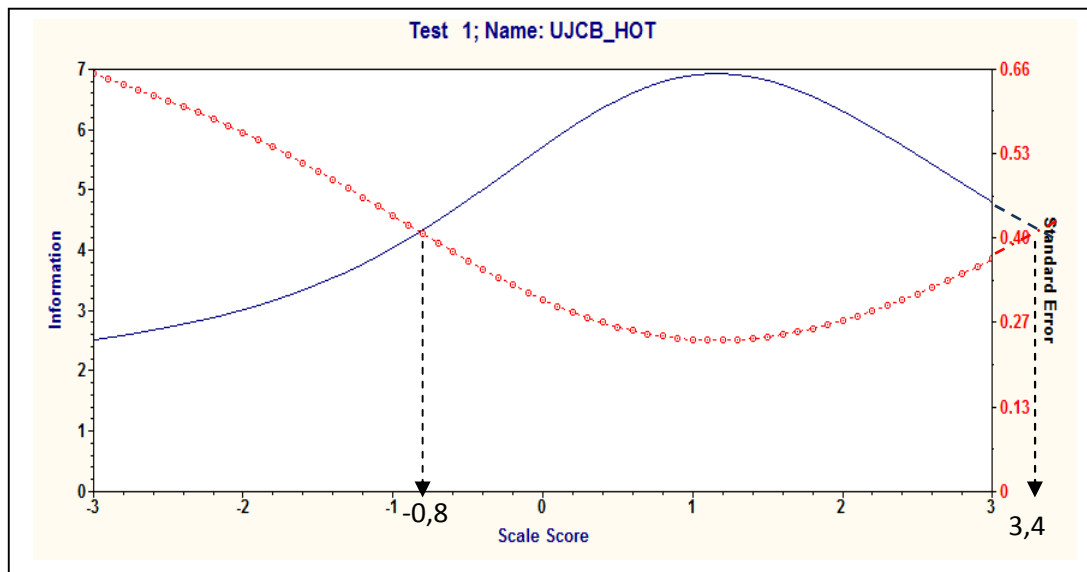
Indeks kesukaran atau tingkat kesukaran (b) untuk skor 0 (b_0), skor 1 (b_1), skor 2 (b_2), skor 3 (b_3), skor 4 (b_4), dan reratanya sebagai *difficulty*. Berdasarkan hasil analisis dengan QUEST, difficulty item-item terletak antara -0,86 sampai dengan 1,06 dengan rata-rata 0 dan simpangan baku 0,42. Item dikatakan baik jika indeks kesukaran lebih dari -2,0 atau kurang dari 2,0 ($-2,0 < b < 2,0$). Jadi berdasarkan difficulty, semua item sebanyak 44 semuanya baik. Tabel 3 dapat diketahui tingkat kesulitan pada masing-masing subaspek dan aspek instrumen untuk masing-masing kategori dalam PCM.

Tabel 3. Tingkat Kesulitan Butir Masing-masing Subaspek untuk Kategori 1, 2, 3, dan 4

No	Aspek	Subaspek	Difficulty	Delta			
				Kategori 1	Kategori 2	Kategori 3	Kategori 4
1	Menganalisis	Membedakan	-0,46	-2,96	0,45	0,92	-0,16
		Mengurutkan	-0,42	-3,09	0,32	1,38	-0,21
		Memberikan ciri khusus	-0,12	-2,69	0,95	1,41	0,00
2	Mengevaluasi	Mengecek	-0,26	-2,00	1,25	0,37	-0,31
		Mengkritik	-0,17	-2,36	1,68	0,19	-0,14
3	Menciptakan	Memunculkan Ide	0,54	0,99	0,32	0,72	-0,23
		Merencanakan	0,53	1,71	-0,28	0,72	-0,16
		Menghasilkan	0,26	-1,10	1,84	0,31	-0,22

5. Fungsi Informasi dan SEM

Hasil analisis dengan Parscale diperoleh fungsi informasi dan *standard error measurement* (SEM). Berdasarkan fungsi informasi dan SEM yang dinyatakan Gambar 2, maka tes ini cocok untuk siswa dengan kemampuan berpikir tingkat tinggi Fisika (θ) dalam kategori sedang, yakni $-0,8 \leq \theta \leq 3,4$.



Gambar 2.
Fungsi Informasi dan SEM pada Kegiatan Ujicoba

Revisi Tes untuk Pengukuran

Berdasarkan hasil uji coba, semua item adalah *fit*. Dua *anchor item* yang tingkat kesulitannya lebih dari satu, yakni item ke-22 (1,03) dan ke-23 (1,06) ditukar dengan item ke-13 (0,31) dan ke-12 (0,38). Penukaran item didasarkan pada kesesuaian sub aspek dan sub materi fisika. Instrumen *PhysTHOTS* terdiri atas dua set tes A dan B yang masing berisi 26 item dengan 8 *anchor item*.

SIMPULAN DAN SARAN

Simpulan

Berdasarkan uraian di atas, simpulan yang dapat ditarik adalah sebagai berikut.

1. Instrumen *PhysTHOTS* dikembangkan pada kemampuan menganalisis, mengevaluasi, dan menciptakan untuk materi fisika gerak, gaya, usaha dan energi, serta momentum dan impuls yang terdiri atas perangkat tes A dan tes B yang masing-masing terdapat 26 item dengan 8 *anchor item*.
2. Instrumen *PhysTHOTS* memiliki karakteristik :
 - a. Instrumen *PhysTHOTS* telah memenuhi validitas isi dengan *expert judgment* dan telah mendapatkan bukti empiris *fit* dengan *Partial Credit Model (PCM)* berdasarkan data politomus empat kategori.
 - b. Seluruh item pada *PhysTHOTS* dalam kriteria baik karena tingkat kesulitannya berada pada rentang antara -2,00 sampai dengan 2,00.

- c. Reliabilitas *PhysTHOTS* telah memenuhi syarat, bahkan termasuk tinggi (koefisien reliabilitas lebih dari 0,90).
- d. Berdasarkan fungsi informasi, *PhysTHOTS* sangat tepat digunakan untuk mengukur kemampuan berpikir tingkat tinggi fisika peserta didik yang berkemampuan dari -0,80 sampai 3,40.

Saran

Berdasarkan karakteristik instrumen, disarankan agar para guru di lapangan melatih peserta didik melalui pembelajaran yang cocok untuk mengembangkan kemampuan berpikir tingkat tinggi untuk mendampingi pembelajaran yang selama ini masih belum optimal dalam mengembangkan kemampuan berpikir tingkat tinggi. Dengan demikian, akan terjadi pengembangan kemampuan berpikir secara utuh dari tingkat rendah sampai tingkat tinggi.

DAFTAR PUSTAKA

- Anderson, L.W., and Krathwohl, D.R. (2001). *A Taxonomy of Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.
- Barnett, J. E and Francis, A.L. (2012). *Using higher order thinking questions to foster critical thinking: a classroom study*. Educational Psychology: An International Journal of Experimental Educational Psychology. <http://www.tandfonline.com/loi/cedp20>. Diakses tanggal 10 Desember 2012
- Bloom, B.S., et al. (1979). *Taxonomy of Educational Objectives: Handbook I Cognitive Domain*. London: Longmans Group Ltd.
- Bond, TG and Fox, CM. (2007). *Applying the Rasch Model. Fundamental Measurement in the Human Sciences (2 rd edition)*. Mahwah, NJ: Lawrence Erlbaum. Bonk
- Brookhart, S. M. (2010). *How to Assess Higher Order Thinking Skills in Your Classroom*. Alexandria: ASCD
- BSNP. (2006). *Standar Kompetensi dan Kompetensi Dasar Mata Pelajaran Fisika Untuk SMA dan MA*. Jakarta: BSNP-Depdiknas
- Chiappetta, E. L and Koballa, T. R. 2010. *Science Instruction in the Middle and Secondary Schools*. New York: Allyn & Bacon
- Depdiknas. (2003). *Undang-undang No. 20 tahun 2003 tentang Sistem Pendidikan Nasional*

- , (2007). *Peraturan Menteri Pendidikan Nasional No 20 Tahun 2007 tentang Standar Penilaian*
- , (2008). *Panduan Penulisan Butir Soal* Jakarta: Direktorat Pembinaan Sekolah Menengah Atas
- Dikpora DIY. (2012). *Hasil Nilai UN 2012 SMA/MA/SMK di DIY* (27 Mei 2012). Yogyakarta: Dikpora DIY http://www.pendidikan-diy.go.id/?view=vberita&id_sub=2692 diakses tanggal 1 Juni 2012
- Djemari Mardapi. (2012). *Pengukuran, penilaian, dan evaluasi pendidikan* . Yogyakarta: Nuha Litera.
- Haladyna, T. M. (1997). *Writing Test Item to Evaluate Higher Order Thinking*. Boston: Allyn and Bacon
- Haladyna, T. M. (2004). *Devoping and Validating Multiple Choise Test Items*. New Jersey: Lawrence Erlbaum Associates, Inc.
- Hambleton dan Swaminathan (1991): *Fundamentals of Item Response Theory*. California: SAGE Publications, Inc
- Linacre, J. M. (2006). *Winstep : Rasch-model Computer Programs*. Chicago: Winsteps.com.
- Linn, R. L. (1999). *Measurement and Evaluation in Teaching* . New York: Macmillan Publishing Company
- Madhuri, G.V. Kantamreddi, V. S.S.N and Goteti, P.L.N.S. (2011). *Promoting higher order thinking skills using inquiry-based learning*. European Journal of Engineering Education. <http://www.tandfonline.com/loi/ceee20>. Diakses 20 Novemver 2012
- Muraki, E., & Bock, R.D. (1997). *Parscale 3: IRT based test scoring and item analysis for graded items and rating scales*. Chicago: Scintific Software Inc.
- Muraki, E. (2003). *Models in PARSCALE*. In M. du Toit (Ed.), *IRT from SSI : BILOGMG, MULTILOG, PARSCALE, TESTFACT*. Chicago: Scientific Software International.
- Nana Sudjana. (1990). *Penilaian Hasil Blajar Mengajar*. Bandung: PT Remaja Rosdakarya
- Oriondo, L.L.and Dallo-Antonio, E.M. (1998). *Evaluation Educational Outcomes*. Manila: Rex Printing Compagny, inc
- Ostini, Remo and Nering, M. L. 2006. *Polytomous item response theory models*. California: Sage Publications, Inc.

- Reedal, K.E. (2010). *Jean Piaget's Cognitive Development Theory in Mathematics Education*. Department of Mathematics and Computer Science – Ripon College. Summation, May 2010, pp. 16-20 <http://ripon.edu/macsum/summation>.
- Ridwan Efendi. (2010). *Kemampuan Fisika Siswa Indonesia dalam TIMSS*. Prosiding Seminar Nasional Fisika 2010 ISBN : 978-979-98010-6-7
- Schraw , G, and Robinson, D.H. (2011). *Assessment of Higher Order Thinking Skills*. New York: Information Age Publishing, Inc
- Sumaji (1990). *Faktor Internal dan Eksternal Siswa Sekolah Menengah Atas dan Kemandiriannya Melakukan Percobaan Fisika*. (Disertasi Doktor). Tidak diterbitkan Yogyakarta: IKIP Yogyakarta
- TIMSS & PIRLS International Study Center. (2012). *TIMSS 2011 international results in science*. Boston: The TIMSS & PIRLS International Study Center, Boston College. Diambil tanggal 5 Januari 2013, dari <http://timss.bc.edu/timss2011/release.html>
- Tsutakawa, R. K., & Johnson, J. C. (1990). The effect of uncertainty of item parameter estimation on ability estimates. *Psychometrika*, 55, 371–390.
- Van der Linden, Wim J and Hambleton, Ronald K. (1997). *Handbook of Modern Item Response Theory* . New York: Springer-Verlag New York, Inc
- Wright, B.D. & Masters, G.N. (1982). *Rating scale analysis*. Chicago: Mesa Press.
- Wu, M., & Adams, R. (2007). *Applying the Rasch model to psychosocial measurement: A practical approach*. Melbourne: Educational Measurement Solutions.